

JourneyDB: A Benchmark for Generative Image Understanding

Year: 2023 | Citations: 160 | Authors: Junting Pan, Keqiang Sun, Yuying Ge, Hao Li, Haodong Duan

Abstract

While recent advancements in vision-language models have had a transformative impact on multi-modal comprehension, the extent to which these models possess the ability to comprehend generated images remains uncertain. Synthetic images, in comparison to real data, encompass a higher level of diversity in terms of both content and style, thereby presenting significant challenges for the models to fully grasp. In light of this challenge, we introduce a comprehensive dataset, referred to as JourneyDB, that caters to the domain of generative images within the context of multi-modal visual understanding. Our meticulously curated dataset comprises 4 million distinct and high-quality generated images, each paired with the corresponding text prompts that were employed in their creation. Furthermore, we additionally introduce an external subset with results of another 22 text-to-image generative models, which makes JourneyDB a comprehensive benchmark for evaluating the comprehension of generated images. On our dataset, we have devised four benchmarks to assess the performance of generated image comprehension in relation to both content and style interpretation. These benchmarks encompass prompt inversion, style retrieval, image captioning, and visual question answering. Lastly, we evaluate the performance of state-of-the-art multi-modal models when applied to the JourneyDB dataset, providing a comprehensive analysis of their strengths and limitations in comprehending generated content. We anticipate that the proposed dataset and benchmarks will facilitate further research in the field of generative content understanding. The dataset is publicly available at <https://journeydb.github.io>.